A Decision Support System for Microbiology Quality Control

Brian R. Jackson, M.D., Joseph D. Schwartzman, M.D.,
Deborah E. Zuaro, M.A.T., M.T. (A.S.C.P.), Edward K. Shultz, M.D.[1]
Department of Pathology and Program in Medical Information Science[1]
Dartmouth-Hitchcock Medical Center
Lebanon, New Hampshire

Abstract

Manual review of antibiotic sensitivity testing results is an essential component of a microbiology laboratory's quality control process. Such review is tedious and prone to human error, however. An expert system is described that remembers which susceptibility patterns are considered typical or atypical by expert reviewers, then uses these to pre-screen future isolates. It uses a similarity function to allow matching against this library when two patterns are close, but not identical. Use of this system allows more efficient and reliable review of the laboratory's antibiotic sensitivity testing results.

INTRODUCTION

One important element of the quality control (QC) process in the microbiology laboratory consists of daily supervisory review of antibiotic susceptibility test results. This allows monitoring of result quality and surveillance of evolving antibiotic resistance.[1,2] Since different species often have limited ranges of susceptibility phenotypes, an organism's antibiogram, or set of susceptibility test results, may also be used to verify the species identification.

This manual review relies on a human's ability to select abnormal patterns from a collection of data. It follows that its effectiveness depends on the ways in which abnormal patterns differ from normal ones, the ability of the reviewer to remember these distinguishing factors, the quantity of data, and the time available for review. Unfortunately, the evolution of antibiotic resistance in bacteria is rapidly expanding the number of "typical" antimicrobial susceptibility patterns.[1]

Computers and knowledge-based systems have long been successfully applied to quality control.[3] Although it is difficult (if not impossible) to fully replace human experts for most of these tasks, knowledge-based systems can improve the ability of humans to perform quality control by doing preliminary screening.[4] Given the continual pressure on the laboratory to cut costs while maintaining or improving quality, computer assistance is becoming increasingly important.

We have developed a system that acquires knowledge each day from an expert (the microbiologist) about which antibiograms are "typical" or "atypical" and then uses these rules to screen subsequent days' results. All antibiograms that are not screened out as typical are then printed out for manual review, along with the most similar typical and atypical antibiograms in the knowledge base for comparison. This allows a more efficient and comprehensive QC review process, reducing the total time required for review and increasing the quality of the information generated. We believe that this tool will enable the microbiologist to more easily pick out laboratory errors, unusual isolates, and trends. The system is currently in daily use, and preliminary evaluation has been positive. Summative evaluation is in progress.

METHODS

Database construction

A relational database was constructed using 4th Dimension (ACI US, Inc.) for the Macintosh (Apple Computer, Cupertino CA). Semantic relationships between different isolate names were defined hierarchically so that the computer could recognize that, for example, an isolate identified as "Gram-negative rod" might potentially be the same as one labeled either *Escherichia coli* or *Pseudomonas aeruginosa*, but that these latter two are definitely distinct.

Data acquisition

A program was written in CCL (Cerner Command Language, Cerner Corp., Kansas City MO) to extract all microbiology results verified on the previous day, along with patient, specimen, and antibiotic therapy information, from our Cerner laboratory information system (LIS) into a text file. This program runs early each morning in batch mode. A second batch program then transfers the file via FTP to a network server. A desktop Macintosh 6400 running Cron (Mark Malson, Hamilton OH) wakes up automatically at 7:00 am each morning and starts up a Hypercard (Apple Computer) program. Hypercard in turn sends a command to Anarchie (Stairways Shareware, Berkeley CA), to transfer the text file from the server. Hypercard then launches 4th Dimension and sends that program a command to upload the data. Hypercard logs any errors that occur

1091—8280/97/$5.00

on the Macintosh end; errors generated by the CCL program or the initial file transfer are monitored by Computer Support personnel. Data acquisition would admittedly be more direct through a Health Level Seven (HL7) LIS interface, were this available to us.

Description of data

Each isolate for which susceptibility testing is performed is typically tested against a panel of between five and eighteen antibiotics. Selection of isolates for susceptibility testing is based on a number of considerations, including body site from which the culture was obtained, relationship to other flora, and species. Furthermore, the antibiotic panels chosen for testing depend on body site and/or species. Approximately 2/3 of these isolates are tested are by microdilution on the Vitek system (BioMerieux Vitek, Hazelwood MO), which gives as output the minimum inhibitory concentration of each antibiotic (MIC). Most of the remainder are tested by Kirby-Bauer (KB) disk diffusion, where the results are the diameters of the zones of inhibition. KB zone sizes are roughly inversely proportional to the log of the MIC for any given species.[5] Some organisms that cannot be reliably tested by either of these methods are tested by ETest (AB Biodisk, Solna Sweden), an agar diffusion method yielding an MIC. Finally, two tests, beta-lactamase expression and high-level aminoglycoside testing to predict beta-lactam synergy, are reported as either positive or negative.

Similarity function

A similarity function was constructed as follows to compare sets of susceptibility test results: First, twenty weeks' worth of susceptibility test quality control (QC) data were entered into a spreadsheet and examined. This represented weekly testing in which nine different reference strains of bacteria are tested with each of the different antibiotic panels, for a total of 169 * 20 = 3380 individual tests. We took the standard deviation (S.D.) of the zone sizes or log MIC's for each individual strain/antibiotic combination (with n=20), assuming that variability in any one test with a given strain could be treated as normally distributed. The standard deviation of each control value was observed to be roughly linearly related to the expected value, so linear regression was employed to estimate the expected standard deviation for a given patient value. This yielded the equations sd = 0.041 * (zone size) + 0.22 and sd = -.0061 * (ln MIC) + 0.057.

Given two isolates tested by the same method, we calculate a distance between their antibiograms by comparing the individual antibiotic tests. This n-space is scaled in each dimension by the expected standard deviation calculated above, allowing a Euclidean distance estimate to be calculated. This estimate assumes that the resistance phenotypes of both isolates are identical, and that differing results are due to analytical variation in the laboratory. Since the number of dimensions involved depends on how many antibiotic tests the two panels have in common, the distance is then normalized by dividing by the square root of the number of antibiotic tests in common. The resulting distance measure can thus be considered to be roughly in terms of standard deviations. (It should be noted however that the absolute value of the distance between two random elements of a normally distributed set is not itself normally distributed.) Beta-lactamase and aminoglycoside synergy results were arbitrarily treated the same as MIC results and assigned the equivalent to a 4-fold dilution difference if discordant.

For simplicity we neglected S.D. variation among different species of bacteria and among different antibiotics. Although controlling for these two factors would refine the metric somewhat, we felt we had insufficient data to do this properly. In addition, previously published work with susceptibility test panels showed that using Euclidean distance, sum of distances, skew Euclidean distance (Euclidean distance on oblique axes based on correlation between variables), and even the correlation coefficient itself as a metric all gave similar clustering of bacterial isolates.[6]

The same twenty weeks' worth of QC data used previously were analyzed by calculating the distances between each test panel/bacterial strain combination. The distances ranged from 0 to 14 between panels performed on the same strain (0 to 4.1 when MIC data was deleted), and from 2.6 to >100 on panels performed on different strains. An arbitrary distance of 2 was chosen as a cutoff for labeling two panels of results as similar. We felt that this would provide reasonable "fuzziness" to the system without risking misclassifying two different susceptibility patterns as similar. When applied to the QC data, this cutoff correctly categorized 68% of the comparisons between panels performed on the same strain and 100% of the comparisons between panels performed on different strains. This improved to 84% of comparisons between identical strains when the MIC data were excluded. The wider variation of the MIC results was due to their highly discontinuous nature, such that each result has only a few possible values, and >80% of the results equal either the highest or the lowest value. We did not feel, however, that this invalidated the use of our metric on MIC results.

It may have been more statistically rigorous to take all clinical isolates for which sensitivity testing was performed over a period of time, perform repeated

testing on these, and then determine the discriminant ability of setting the metric at different points. However, this would have expended significant resources for what we considered would be a small gain.

Knowledge Acquisition

All of the first day's susceptibility results were reviewed by a microbiologist, and all of the isolates were classified as either "typical" or "atypical" antibiograms. These were then stored with the appropriate designations in a separate file of the database. Since then, all results have first been screened against this library. Any set of results matching (i.e. distance < 2) a panel stored as "typical" is so labeled in the daily QC report, while any matching an "atypical" isolate is flagged and printed along with the previously stored atypical pattern. Isolates matching no stored isolates are labeled as indeterminate, and the closest isolates from both the typical and atypical libraries are printed for comparison. Indeterminate patterns are then classified by a microbiologist as either typical or atypical and added to the knowledge base. (Atypical patterns are edited down to include only those antibiotic results that define the pattern as atypical.) In this way, knowledge acquisition can occur semi-automatically as a consequence of daily result review.

For an isolate to match an antibiogram in the comparison library, three conditions must be met in addition to the distance between their patterns. First, the two must be potentially the same species as determined by their semantic relationship. Second, the test methods (KB or MIC) must match. Third, the antibiotics in that isolate's antibiogram must either include all antibiotics in the comparison atypical antibiogram or be a subset of the antibiotics in the comparison typical antibiogram. Conceptually, this follows from the principle that for an antibiogram to be considered typical, each component must also be typical. Conversely, any atypical component will define the entire pattern as atypical.

In the first phase of development, results of the daily review have been entered into the database by the programmer (B.J.) This process takes approximately three minutes each day. Once the system has sufficiently matured, the microbiologists will take over the task of knowledge entry. Our eventual goal is for the entire process to be paperless, with simultaneous interactive result review and knowledge entry on the computer.

Also during this first phase of development, all manual review is being performed in duplicate to ensure the accuracy and reproducibility of the classifications being entered into the knowledge base. This, combined with the fact that a number of changes have taken place in the laboratory over the past 6 months, has decreased the frequency of knowledge acquisition from daily to rather sporadic. This should become more regular in the future, however, particularly as the knowledge base grows and leaves fewer antibiograms needing to be classified each day. We expect that over the long run, the time required for result review on the computer will more than be made up for by the time savings of not having to review so many susceptibility results each day. Thus, even if this system did not improve the quality of daily QC review, which we believe it does, it will pay for itself in terms of time saved.

System Evaluation

The LIS report which had been used for review of susceptibility testing was replaced by one which instead used the 4th Dimension database. This contained the same data in the same format as the previous report, except that the label "typical," "atypical," or "unclassified" (if there is no match found in the knowledge base) was added to each susceptibility panel, as well as the closest comparison antibiograms and the distances to these. Each day, one antibiogram is randomly selected to be labeled "unclassified" regardless of whether it matches a typical pattern or not (although if it matches an atypical pattern, it is labeled as such). The reviewers are blinded to when this occurs. This ensures periodic review of the typical patterns in the knowledge base, and creates a means of testing the reproducibility of the reviewers' decisions. (Each atypical pattern is brought up for review each time a match is made to it, obviating the need for further review of these.)

We are conducting a crossover trial, in which two microbiologists review the daily susceptibility reports. One reviews the report as just described, and the other reviews that report with the screening information removed. Halfway through the trial, the two reviewers will trade report types. Addition of patterns to the knowledge base has been suspended until the end of the trial, and both reviewers have available a printout of all of the typical and atypical patterns in the knowledge base.

Primary outcome measures are time required for daily review and accuracy in identifying patterns as typical or atypical, using the consensus decisions of the microbiologists as the gold standard. In addition, internal consistency of the classification will be evaluated by means of the blinded review of typical patterns above, as well as by performing retrospective nearest-neighbor analysis on the entire set of observed antibiograms.

RESULTS

Laboratory volume

Our laboratory, which serves a moderate-sized academic medical center as well as several outpatient clinics, processes an average of 149 cultures per day. From these, sensitivity testing is performed on an average of 17 isolates per day. The 3064 isolates tested since January encompass 94 different species, of which the 7 most common species represent 78% of the total test volume.

Database reliability

After 180 days of continuous operation, the database contains 104 MB of data. The time required to import each day's results averages 18 minutes, and is very gradually increasing as the database grows. This is due to the searching that takes place to prevent duplicate entries. The automatic file transfer and import routines have been generally reliable, although occasional errors occur due to delays in the LIS batch stream. Such errors occur approximately twice per month, and daily maintenance is therefore necessary.

Classification and screening

Up to now, twelve days worth of antibiograms have been manually reviewed and classified. This has yielded 121 typical patterns and 10 atypical patterns. 13 additional atypical patterns have been added, not in response to actual isolates observed in the laboratory, but based on phenotypes that the microbiologists would like flagged should they occur. Four atypical patterns were later reconsidered and deleted, leaving a current total of 19. These 140 patterns are sufficient to screen out slightly less than half of new antibiograms. In the most recent week, for example, there were 113 antibiograms generated by the laboratory. Of these, our system categorized 46 as typical, 6 as atypical, and 61 as unclassified.

Data validation

The first 91 antibiograms to be classified were tested retrospectively against the knowledge base to see if their classifications would remain the same. There was a discrepancy in a single case, where a coagulase-negative *Staphylococcus* sensitive to penicillin was classified as typical early in the week, and then later in the week one with an identical susceptibility pattern was classified as atypical. The case was discussed with the microbiologist, who had in fact changed his mind during that time about whether he wanted this particular pattern flagged as atypical or not. In practice, the system will handle such changes gracefully, since matching to an atypical pattern takes precedence over matching to a typical one. Additionally, if any clinical isolate matches both an atypical and a typical pattern, both of these will be printed for comparison, and the microbiologist may then choose to delete one of the two.

Since then, an additional 15 atypical and 10 typical classifications have been reviewed. Both microbiologists agreed with all of the typical classifications. The atypical patterns fell into three categories: The microbiologists agreed with 7 of the classifications, all of which involved uncommon, but not unheard of, resistance patterns that the microbiologists nonetheless wished brought to their attention. They disagreed with 7 others which involved patterns they had previously wanted brought to their attention, but had since changed their minds. These were therefore deleted from the library of atypical patterns. Finally, they disagreed with one classification which involved an isolate identified at the time as "gram negative bacillus" and matched atypical patterns for Proteus mirabilis and Enterobacter species (the isolate turned out to be neither of these). This represented a false-positive.

DISCUSSION

The QC review system we have described in this paper can be thought of as a sort of accessory memory for the microbiologist. As such, it constitutes a simple yet effective expert system employing pattern recognition based on supervised learning and modified nearest neighbor analysis.[7] Other approaches were considered and rejected based on the nature of both the problem and the data involved. For example, although inductive reasoning systems have proved useful in many situations, they tend to perform best in situations where the numbers of inputs and outputs (i.e. observations and diagnoses) are small compared to the number of example cases.[4] Our situation was just the opposite.

Another, more traditional approach to expert systems has been to parse and symbolically represent the logic used by the expert to arrive at a diagnosis. This sort of knowledge acquisition has been described by some as the most difficult element of expert system development.[8] We have taken a simpler approach, acquiring primarily the end decisions of the expert rather than the logic used to arrive at those decisions. The user is then free to supplement the knowledge base with additional rules as he/she deems appropriate. This limits the system's "intelligence," but increases simplicity and speed of knowledge acquisition.

We could have designed our system to take a more active role in classifying patterns and automatically incorporating them into the knowledge base. By doing so, however, we would have risked losing the self-knowledge that is essential for reliable expert system performance.[9] We thought it therefore

desirable that all antibiograms that are not extremely close to a known classified pattern be manually reviewed by the microbiologist before being included in the knowledge base.

All expert systems face the challenge of keeping the knowledge base up to date.[10] This is particularly important to our application due to the rate of evolving antibiotic resistance. By incorporating knowledge acquisition into the existing QC review process, we have made it manageable, relevant, and responsive to the changing epidemiology of the hospital environment.

Ours is not the first expert system designed to review microbiology results. An important successful example is the GermWatcher system at Washington University, which uses Centers for Disease Control (CDC) criteria to review culture results and determine which are likely to represent nosocomial pathogens.[11] Another group has used cluster analysis of susceptibility tests to determine the likelihood that same-species isolates from different patients are the same strain and thus potentially represent cross-infection.[6] The uniqueness of our approach lies in the focus on the microbiology laboratory itself and the mechanism for continuous knowledge acquisition and revision.

Some limitations should be acknowledged. First of all, like all expert systems, ours is limited by the quality of the knowledge entered into it. The fact that the expert him/herself is doing the knowledge entry has some advantages as mentioned above, but may make the knowledge somewhat less reliable than if it were entered with the assistance of an actual knowledge engineer.[8] We believe the system of blindly rechecking isolates will help control for this, however. Second, this system is currently only useful for retrospective review of results. If the knowledge were available in real time, so that a technologist entering results could receive immediate feedback every time an atypical pattern was entered, additional benefits could be realized. This would require either redesigning the system to run within the LIS itself or interfacing with the LIS in a way that we are presently unable to do. Third, although the system was initially conceived to take advantage of results review already taking place, it was later decided to perform more time-consuming duplicate review as a means of ensuring the quality of the knowledge being entered into the system. In the sense that this provides added validation, this is a good thing. It does have the unfortunate effect, however, of making the system less convenient for the microbiologists, and so knowledge acquisition has been less frequent than the daily capture initially anticipated.

CONCLUSION

We have developed a quality control decision support tool which incorporates many features of traditional expert systems, while avoiding some of the difficulties often associated with them. Further, this was accomplished with relatively little development time and effort. We feel that this system will improve our quality control process while saving time and money. In addition, it may help the laboratory to more effectively monitor emerging resistance in the hospital.

References

1. Hindler J. Non-traditional approaches for quality control of antimicrobial susceptibility tests. Adv Exp Med Biol 1994;349:67-85.
2. Sahm DF, O'Brien TF. Detection and surveillance of antimicrobial resistance. Trends Microbiol 1994;2:366-371.
3. Nusbaum NJ. Computers in the clinical laboratory. Medical Hypotheses 1995;44:70-72.
4. Spackman KA. Quality assurance, knowledge-based systems, and machine learning. Med Decis Making 1991;11:153.
5. Acar JF, Goldstein FW. Disk susceptibility test. In: Lorian V, editor. Antibiotics in Laboratory Medicine. 4th ed. Baltimore: Williams & Wilkins; 1996. p. 1-51.
6. Giacca M, Menzo S, Trojan S, Monti-Bragadin C. Cluster analysis of antibiotic susceptibility patterns of clinical isolates as a tool in nosocomial infection surveillance. Eur J Epidemiol 1987;3(2):155-163.
7. Duda RO, Hart PE. Pattern Classification and Scene Analysis. New York: John Wiley & Sons; 1973.
8. Edwards M, Cooley RE. Expertise in expert systems: knowledge acquisition for biological expert systems. Comput Appl Biosci 1993;9(6):657-65.
9. Sumner W, Shultz EK. Expert systems and expert behavior. J Med Sys 1992;16(5):183-193.
10. Giuse DA, Giuse NB, Miller RA. Evaluation of long-term maintenance of a large medical knowledge base. J Am Med Inform Assoc 1995;2:297-306.
11. Kahn MG, Steib SA, Fraser VJ, Dunagan WC. An expert system for culture-based infection-control surveillance. SCAMC Proc. 1993;171-5.